DOCUMENT RESUME

ED 260 218 CE 042 053

AUTHOR Parker, Barbara L.
TITLE Summative Evaluation in Training and Development: A Review and Critique of the Literature, 1980 and 1983. Training and Development Research Center: Project Number Four.
INSTITUTION Minnesota Univ., St. Paul. Dept. of Vocational and Technical Education.
PUB DATE Oct 84
NOTE 58p.; For related documents, see CE 042 050-057.
PUB TYPE Information Analyses (070) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Evaluation Methods; *Evaluation Needs; Evaluation Utilization; Evaluators; *Job Training; *Labor Force Development; Literature Reviews; *Program Evaluation; *Summative Evaluation

ABSTRACT

From 1980 through 1983, 41 articles in selected professional journals dealt with summative evaluation of training and development efforts in business and industry. All of the standard components of program evaluation were addressed either directly or indirectly in them, in varying degree and depth. Although one must contact the primary audience to determine the purpose of an evaluation, only 26 articles mentioned this step. While the majority of articles discussed evaluation questions, few linked them to specific audiences. The purpose seemed to be to prove to funders that the training function is worthwhile. Possible constraints on the evaluation process were mentioned in 21 articles. The frameworks implied in these articles were clustered around decision making, goal achievement, and accountability. Eight possible designs/strategies were mentioned: experimental study, quasi-experimental study, survey, personnel or client assessment, systematic "expert" judgment, and case study. Tactics, both technical and managerial, were also mentioned. Thirteen articles referred to the reporting component of evaluation. A third mentioned planning, the references ranging from single statements to a matrix guide. Only 37 percent referred to the evaluator's role and/or qualifications. (An appended master data chart lists the articles and summarizes the information found in each.) (YLB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED260218

TRAINING AND DEVELOPMENT RESEARCH CENTER

Project Number Four

SUMMATIVE EVALUATION IN TRAINING AND DEVELOPMENT: A REVIEW AND CRITIQUE OF THE LITERATURE, 1980 THROUGH 1983

Barbara L. Parker
October 1984

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

university of minnesota
DEPARTMENT OF VOCATIONAL AND TECHNICAL EDUCATION • ST. PAUL, MINNESOTA

CE042053

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R Swanson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Preface

Objective evaluation of training and development programs is neither widely practiced nor accepted. The common explanations for this condition range from indictments of the lack of professional competence in evaluation to a philosophical position that we can feel and see truth and do not need to reason an organization into it.

Barbara L. Parker took on the task of reviewing the 1980-83 training and development summative evaluation literature in order that we better understand contemporary practices in the profession. Her monograph represents one of the most thorough and thoughtful documents on the topic of evaluation practices in training and development. It should serve us well in our understanding of good evaluation practices, actual practices in the profession, and pinpointing efforts at improving evaluation practices on a program by program basis.

Richard A. Swanson, Director
Training and Development Research Center
University of Minnesota
1954 Buford Avenue
St. Paul, MN 55108

Summative Evaluation in Training and Development: A Review and Critique of the Literature, 1980 Through 1983

The topic of impact, or summative, evaluation is a concern of many training specialists. According to a recent questionnaire survey, summative impact evaluation questions were given as the most critical questions facing the corporate trainers who responded (Olivas & Inman, 1983).

The evaluation of training and development efforts in business and industry has been reported for at least the past 30 years (Lindholm, 1953). In 1960, Kirkpatrick described techniques for assessing learner reaction, learning, behavior, and results of training with respect to organizational goals. In addition, he discussed the problems involved with such areas of evaluation and cited a number of studies as examples of possible evaluation procedures. Kirkpatrick stressed the importance of having evaluation skills and using evaluation results in the training and development field.

In 1968, a sample of U.S. and Canadian industries were surveyed as to the extent to which Kirkpatrick's above four levels of evaluation were being conducted. It was found that most evaluation efforts were conducted by assessing learner reaction to training. Few attempts were made to measure results in terms of company objectives, a domain the authors admitted is considerably more difficult to assess, but most desirable (Catalanello & Kirkpatrick, 1968).

An in-house survey conducted in the late 70's of 44 courses at New England Telephone revealed that students' reaction data had been summarized for 85 percent of the courses, but only 28 percent had been evaluated in terms of post-training job performance and only 10 percent in terms of impact on organizational results (Smith, 1980a).

As recently as 1979, a study done by Smeltzer of 285 companies revealed the same results reported by Catalanello and Kirkpatrick 10 years earlier: The majority of the companies evaluated the reactions of the trainees and their opinions regarding

"whether or not they have learned anything," immediately upon completion of the training.

The purpose of this monograph is to review and analyze the literature from 1980 through 1983 dealing with summative (or impact) evaluation of training and development efforts in business and industry. The 41 relevant articles were located through a manual search of professional journals in the training and development field. An analysis is made to determine to what extent evaluation process components used by professional program evaluators are addressed by each of the articles reviewed. The program evaluation components of audience(s), question(s), purpose, context/constraints, methods and procedures, reporting, and planning (Brinkerhoff, Brethower, Hluchyj & Nowakowki, 1983; Welch, Note 1) are summarized in Table 1 and described in more detail in the body of the text which follows.

TABLE 1

PROFESSIONAL PROGRAM EVALUATION COMPONENTS AND DEFINITIONS

Audience(s)	Person(s) for whom the evaluation is being conducted.
Question(s)	Audience concerns regarding the program stated in question form and having action implications.
Purpose	General reason for gathering information or evidence; usually in order to contribute to some form of decision making.
Context and Constraints	The social, psychological, and political milieu in which the program exists, including factors which influence the way the evaluation can be carried out.
Methods and Procedures • **Frameworks/Models** • **Strategies/Designs** • **Tactics: Technical and Managerial**	Ways of gathering, analyzing, and interpreting the evidence to answer the evaluation questions.
Reporting	Written and/or oral presentation of evaluation process and findings.
Planning	Designing the evaluation prior to implementation.

In addition to the components, the concomitants of utilization of results and the evaluator's role and qualifications are investigated and discussed. The Master Data Chart in the Appendix lists the 41 reviewed articles and summarizes the information found in each.

Audience(s)

Every evaluation has a primary audience--the person for whom the evaluation is being performed. Usually, the primary audience is the primary decision maker with regards to the program being evaluated. The importance of identifying the audience lies with the ultimate utilization of the evaluation results.

Twenty-six of the 41 articles (63%) referred to the concept of an audience for the evaluation. Brinkerhoff (1981) states, "Above all, evaluation should be a useful activity, responsive to the needs and/or interests of particular audiences" (p. 67). Putman (1980) also stresses the identification of the audience(s): "Vague words like 'management' are rarely adequate here--which specific persons in management will be the decision makers? The more specific your answer to this question, the more focused and impactful your results can be" (p. 38). Coffman (1980) refers to the primary audience as the "key client" (p. 85).

In order to determine the purpose of an evaluation (a component discussed later), one must contact the primary audience. Of the articles reviewed, just slightly over half mention this important step.

Question(s)

Many articles reviewed here state that the "what" of evaluation should be whether or not the training objectives have been met. But simply referring to those objectives is not necessarily the best or only source of what to evaluate.

When an external evaluator evaluates the effectiveness of a program, he or she usually ascertains what the concerns of the primary audience are regarding the program by way of an interview. (The concerns are not necessarily negative.) Evaluation

questions are then developed out of these audience concerns. For any particular training program, then, there could be different evaluation questions, depending on the primary audience's concerns.

The evaluation questions drive the evaluation. They need to be on-target with the primary audience's concerns. They need to be specific and have some action implication as a result of the answers or evidence eventually provided (Welch, Note 1). Smith (1980a) puts it this way: "Planning the evaluation around specific questions or decisions increases the likelihood that the findings will produce concrete actions" (p. 74).

While the majority of articles reviewed do discuss "what should be evaluated," few derive these "whats" from a process of determining a particular audience's concerns. Putman (1980) is one of the exceptions. He makes this point: "Once you know by whom your data will be used, find out what kind of data they actually find useful" (p. 39). Problems in this area can be avoided by identifying and involving primary decision makers from the planning phase on, thereby preventing the possibility of conducting an evaluation based on wrong or insufficient concerns and questions--the "error of the third kind" (Mitroff & Featheringham, 1974).

Purpose

There are a number of possible purposes or reasons for conducting a program evaluation (Anderson & Ball, 1978). Table 2 lists those general ones which are given either explicity or implicitly in these 41 articles. (Since several articles simply discuss the general topic of evaluation instead of describing an actual evaluation, they may allude to more than one reason for conducting an evaluation.)

TABLE 2

PROGRAM EVALUATION PURPOSES
REFERRED TO BY ARTICLES IN THIS REVIEW

	Purpose	Number of Articles
#1	To contribute to decisions regarding program installation.	6
#2	To contribute to decisions regarding program modification.	16
#3	To contribute to decisions regarding program continuation, expansion, or contraction.	37
#4	To obtain evidence to rally support.	13

As shown in Table 2, the general purpose of training and development evaluation which is at least implied in 90% of these articles is to provide information for decisions about continuation, expansion, or contraction of training. The summative evaluation objective in these articles seems to be to prove to those who hold the purse strings that the training function is worthwhile.

It appears that senior management (from whose ranks often comes the primary audience), is demanding accountability for its training and development expenditures and that this pressure is necessarily shifting the concern, and thus the purpose of much evaluation, from assessing at the learner reaction level (and even the learning and behavior levels) to Kirkpatrick's level of bottom line results of training. Brinkerhoff (1981), however, decries this notion that results of training are now considered the only "right" thing to evaluate. He states, "The rightness or wrongness of what is evaluated should be determined by the purpose for conducting the evaluation" (p. 67). To reiterate, this purpose should reflect the audience's concerns.

Thirteen of the 41 articles (32%) indicated that evaluation should be done in

order to gain credibility for the training and development function (purpose #4)--to obtain management's favor, if you will. The authors of two of these articles viewed evaluation as one way to avoid or reduce skepticism of the training and development function (Cook, 1980; King & Roth, 1981). This latter purpose for evaluation is brought out by a minority of the articles reviewed, however. The topic of impact evaluation is receiving a lot of attention, in the literature, but the consensus among most of the authors is that it is seldom actually carried out.

Context and Constraints

Before looking at the "how" of evaluation, it is appropriate to address the importance of considering the context and constraints which surround the evaluation process. These two factors are often interrelated.

Knowing and understanding the context in which an evaluation is to be conducted is crucial. An evaluator must be/become acquainted with the milieu in which the program has taken place and within which he or she will be operating. This milieu involves all the psycho-social-political processes which interact and impinge on the program and on the evaluation in one way or another (Anderson & Ball, 1978). This is where the interpersonal skills of the evaluator are critical (Brandenburg, 1982b). Chris Argyris (1962) describes the evaluator's challenge when he writes:

> It has been shown that interpersonal openness, experimentation, and trust tend to be inhibited in organizations. The same may be said for the concern for truth for its own sake. The sharing of knowledge is not a living value since that could lead to one's organizational survival being threatened. (p. 192)

He also states that the political climate of an organization may prevent openness and trust from ever forming, no matter how good the interpersonal skills of the evaluator, with the entire evaluation effort being seriously jeopardized even

before it begins.

The articles by Smith (1980a) and Putman (1980) also mention this potential problem. For example, instructors or trainees may suspect that they, rather than the training, are being evaluated, and may not cooperate fully (Smith, p. 74).

When an external evaluator is involved, someone who knows how to get things done in the company should act as liaison for the evaluator (Smith, 1980a, p. 75). The need is to be aware of and understand the context in which an evaluation is to be conducted. However, most evaluative efforts are geared only to the immediate aspects of training and do not reflect the context of the total organization (Brandenburg, 1982a, p. 18).

Contextual constraints are often addressed in the literature, but those which affect evaluation need to be separated from those which impinge on training. For example, contextual constraints such as the influence of the organization and the nature of the manager to be trained not only should be addressed during summative evaluation, but also during the needs assessment, course design phase, and the development of training (Clement & Aranda, 1982).

Possible constraints on the evaluation process are mentioned in 21 of the articles reviewed. Such factors as "day-to-day demands of the work environment" and "the vast number of uncontrollable variables present in real life" (Blakeslee, 1982; Dobbs, 1980; Hawes, Hutchens & Crittenden, 1982; Mezoff, 1981), "political forces", "personal values of the decision-maker" (Brinkerhoff, 1981, p. 69), and regulations or labor contracts (Putman, 1980). While these may be difficult or impossible to manipulate, they need to be identified and dealt with as effectively as possible.

Brown (1982) discusses constraints on using the more effective evaluation techniques such as direct observation of a trainee following training: "Direct observation of behavior is costly, time consuming, and for most businesses and industries,

just not practical!" (p. 13). (Brown goes on to point out a constraint on attempting to measure management behavior in the first place: much of it, such as problem solving, is covert and cannot be measured directly.)

Many "constraints", however, may be self-imposed. The following causes of evaluation problems have been offered: knowledge/skill deficiencies of the evaluators, poor or incomplete course documentation, difficulty developing valid and reliable designs and instruments, everchanging course content, defensiveness of instructors or trainers during an evaluation, lack of higher management interest in or support of evaluation, expense, an environment which limits application of findings, the long time lag needed to show effects, and lack of access to personnel and availability of various sources of data (Blakeslee, 1982; Cornwell, 1980b; Mezoff, 1981; Putman, 1980; Salinger & Deming, 1982; Smith, 1980a). The constraint presented in Cook's (1980) article as given by participants at a seminar on training evaluation seems to wrap up the above list well: "Many trainers simply don't know how to do such studies" (p. 5).

Most of these latter constraints can often be controlled and/or prevented with a planned, systematic approach to evaluation, along with trainers who possess evaluation competencies. Training program evaluation doesn't have to be difficult, costly and time-consuming (Dobbs, 1980; Grenough and Dixon, 1982). Zenger and Hargis (1982) argue that their experience confirms the following: (a) the tools for adequate research techniques exist and are not that difficult to construct, (b) management usually supports assessment, (c) research is not expensive compared to total training expenditures, and (d) good research design can be followed (p. 16).

Methods and Procedures

The methods and procedures component of evaluation deals with gathering, analyzing, and interpreting the information or evidence for answering the evaluation questions. The preliminary steps of identifying the primary audience and stating

the right questions (and thus defining the evaluation's purpose), are important determinants for designing appropriate evaluation methods and procedures (Anderson & Ball, 1978). Few of the articles mention the process of linking these two components of audience and evaluation question(s). Instead, most of the authors move right into discussing what they seem to believe should be studied in any evaluation. This increases the possibility of being off-target from the beginning and thereby using inappropriate methods which may result in an ineffective evaluation effort.

Within this component of methods and procedures are the sub-components of framework or model, strategies/designs, and tactics.

Framework or Model

The classic framework or model of evaluation in the field of training and development is Kirkpatrick's four-step approach (Catalanello & Kirkpatrick, 1968; Kirkpatrick, 1960). Of the articles reviewed, one presented an additional model for practitioners today (Bakken & Bernstein, 1982).

Table 3 lists 10 general frameworks or models of evaluation which are familiar to those in the program evaluation field (Welch, Note 1).

Training and development specialists who seek to evaluate their programs in a systematic way would be aided by a knowledge of these various orientations. Such models help to organize mentally the evaluation project and tie appropriate methods to the intended purpose. Except for Bakken and Bernstein (1982) and the authors who refer to Kirkpatrick's (1960) model (Bakken & Bernstein, 1982; Connolly, 1983; Dopyera & Pitone, 1983; McAfee, 1982; Smith, 1980a, 1980b; Zenger & Hargin, 1982), it isn't known how many of the other authors use any conceptual framework for evaluation. The frameworks implied in these articles are clustered around Decision making, Goal Achievement, and Accountability. (Overlap exists since many articles refer to or imply more than one purpose for evaluation.)

Goal Achievement appears to be the orientation used in most of the articles

Table 3

General Frameworks/Models of Program Evaluation (Welch, Note 1) and Percentage of Articles Addressing Each

	FRAMEWORK/MODEL	ADVOCATE	KEY EMPHASIS	IMPLICATIONS	PERCENTAGE OF ARTICLES
1.	Curriculum development	Cronbach, Welch & Walberg	Improving curricula.	Student testing, teaching analysis, controlled situation.	0%
2.	Accreditation	Professional Associations	Certification.	Site visit committees, self study.	0%
3.	Consumer protection (information)	Scriven	Effects on users.	Avoid co-option, skeptic mind-set.	0%
4.	Description (portrayal)	Stake, Hamilton, Parlett	Understand activities and values.	Fieldwork, observation, case studies.	0%
5.	Decision making	Stufflebeam, Welch, Alkin	Providing information to aid in decision making.	Define decision needs, provide information.	24%
6.	Discrepancy analysis	Provus	Comparing performance with standards.	Defining standards, evaluator is a team member.	0%
7.	Goal achievement	Popham, Tyler	Achieving program goals.	Clear goal statements.	68%
8.	Accountability	Lessinger	Cost effectiveness.	Need program costs, benefits clearly measured.	46%
9.	Social change	Weiss, Patton, Guttentag	Bring about social reform.	Action alternatives defined.	0%
10.	Instructional research	Tyler, Walberg	Understanding instructional process.	Focus on teachers and teaching.	0%

(68%). Accountability is referred to next most often (46%), while a Decision making framework was alluded to in 24% of the articles. Of those articles having a Goal Achievement orientation to evaluation, some focus on training program goals while others emphasize company or organization goals. This orientation is quite straight-forward: Were the goals accomplished or not? A disadvantage of this framework is its narrow focus: unintended outcomes and side effects may be ignored. The Accountability orientation is also quite clear-cut, though not necessarily simple to perform: Was the training worth the cost?

The Decision Making orientation, however, is more open-ended and lends itself to more flexibility than the above two frameworks. Once the decision needs are defined, the task is to provide the information to aid in decision making. The decision needs may in fact be to ascertain if goals were achieved or if the training was cost effective. Bakken and Bernstein (1982) combine decision making and goal achievement in their model.

Summative and formative evaluation are often interrelated in using a decision making framework. Upon gathering information, there are usually two basic decisions. The first is to continue or to discontinue the program (a summative evaluation decision). If to continue, the second is to decide which aspects of the training program could be improved upon (a formative evaluation decision) (Smith, 1980a).

Unless the evaluator is also the decision maker (more often the case in formative evaluation), the required emphasis is on obtaining valid information for someone outside of the training function to use in making rational decisions about the training's worth and value.

There is no one correct framework or model. However, awareness of various models and how they can help one conceptualize and plan an evaluation aids training and development specialists who plan, perform, and/or assist with evaluation.

Strategies/Designs

As stated earlier, the evaluation questions direct the evaluation process because they dictate the evidence needed which in turn determines the design(s) or strategies to be used. The evaluator needs to ask what the best way is to go about obtaining the needed information (evidence), considering the constraints of time, cost, political climate, and so forth. A different strategy or strategies may be required for each separate evaluation question (Welch, Note 1). Two of the reviewed articles (Connolly, 1983; Salinger & Deming, 1982) picked up on this concept.

Table 4 lists eight possible designs or strategies used by professional program evaluators (Anderson & Ball, 1978), along with data indicating which ones are referred to in these reviewed articles. Reference to a design/strategy by the authors may range from alluding to it without actually naming it, to giving a detailed description.

TABLE 4

FREQUENCY OF REFERRAL TO EIGHT EVALUATION DESIGNS/STRATEGIES BY THE 41 REVIEWED ARTICLES

	Design/Strategy	No. of Articles	Percent of Articles
#1	Experimental Study	4	10
#2	Quasi-experimental Study	15	37
#3	Survey	33	80
#4	Personnel or Client Assessment	10	24
#5	Systematic "Expert" Judgment	2	5
#6	Case Study	4	10

Before continuing, some similarities and differences between evaluation and research should be mentioned. Accuracy in information gathering and analysis is

stressed in both research and evaluation, using technical and ethical procedures. However, evaluation is usually decision-oriented, while research is strictly knowledge-oriented (Welch, Note 1). This seems to be the dichotomy which Putman (1980) discusses, but he dismisses rigorous experimental design instead of seeing it as one of several possible evaluation designs. He offers an interesting perspective on this topic: "Highly 'rigorous' methodology and design are typically not appropriate in HRD evaluation; using them is like using a microtome to slice cheese. Great tool--wrong job" (p. 38).

#1 - Experimental study. A few of the reviewed articles stress the need for rigorous, "tight", experimental designs (Brown, 1982; Zenger & Hargis, 1982), while others stress practicality (Brinkerhoff, 1981, 1983; Hahne, 1981; Putman, 1980; Salinger & Deming, 1982).

According to Table 4, four of the 41 articles mention experimental designs (in which pre-post measures of both a treatment and control group are taken and all group members have been selected randomly) (Training Today, 1980; Smith, 1980b; Zenger & Hargis, 1982).

It is very difficult to carry out valid, highly "rigorous" experimental research (Design/Strategy #1) outside of a laboratory setting, e.g., in the "real world". Experiments which study a single variable are geared for the physical sciences, not the behavioral sciences. The internal validity of the experiment is threatened by innumerable extraneous variables, few of which can be kept constant. When attempts are made to control variables other than the treatment, a contrived situation may result which resembles little of the dynamic and often unpredictable events within human environments (Weiss & Rein, 1970). More than a decade ago, Campbell (1971) and Argyris (1968) questioned the use of rigorous research in attempting to understand and evaluate organizational activities better.

In the quest for "scientific measures", people in organizations tend to overlook these facts. Rigor is sought at the expense of reality. When reality is compro-

mised, so is the validity of results from such studies. Thus, an evaluation should seek information about the effectiveness of a program given its particular milieu, not by denying it.

#2 - Quasi-experimental study. When members of the treatment and control groups are not selected randomly, the research design is called "quasi-experimental" (Campbell & Stanley, 1966). Such designs are referred to in 37% of the articles reviewed. While being less rigorous than experimental designs, they still attempt to control for some, at least, of the extraneous variables. Quasi-experimental designs are helpful in evaluating at the levels of learning and behavior (Kirkpatrick's model), especially where evidence of an observable and valid pre-post training change is sought.

In this literature, pre-post measures with a control group have been obtained using test results (Hanson, 1982), questionnaires, observations, interviews, and analysis of work-reporting data (Cornwell, 1980b), including performance appraisals (McAfee, 1982).

In addition to a control group, Powers (1983) proposes a third group which receives an alternate treatment than the experimental group treatment. This procedure would determine, for example, if similar results could be obtained with an entirely different type of training.

A "multiple baseline" design is one in which two or more sets of baseline data are collected across behavior or performance variables, across subjects or groups, or across settings or locations. Successful intervention (training) should affect only one of the baselines (that of the trained group), with the other(s) acting as controls (Brown, 1980; Powers, 1983).

Some authors in the training and development literature discuss the "costs" of using a control group such as the difficulty in setting up matched groups, holding environmental variables constant during the study and ethical/economic problems

of withholding training from employees who may benefit from it (Brethower & Rummler, 1979; Brown, 1980).

A pre-post test design without a control group assumes that changes between the pretest and posttest were brought about by the treatment (training). There are many extraneous variables besides training in the workplace which can bring about change, such as experience or some change in the organization. The use of pre-post test design with no control group is recommended in research only under certain circumstances: "when the dependent variable is reasonably stable, when the interval between the pretest and the posttest can be kept short, and when it is impossible to obtain a control group" (Borg & Gall, 1979, p. 540).

Given these recommendations for researchers, one can see the possible inadvisability of foregoing a control group when using pre-post measures for summative program evaluations. Immediate posttest scores (those obtained right after training) say nothing about carry-over and/or maintenance of the behavior on the job; the stability of newly-learned job skills is essentially unknown (especially for less repetitive, non-technical skills), and in most instances, it should not be impossible to obtain a control group.

Time-series analysis is a quasi-experimental design which doesn't use a control group. Data is collected on a single group of subjects at periodic intervals and the treatment (training) is administered between two of these intervals. However, for measuring long-term effects, an extended time-series analysis with a comparison group is recommended (Bakken & Bernstein, 1982).

#3 - Survey. The survey strategy is a systematic method for obtaining standardized information from all respondents in a particular sample. In other words, the same instrument must be administered to all subjects under conditions which are as similar as possible (Borg & Gall, 1979). Survey data may be collected in several ways: through questionnaires and/or interviews (self-reports and reports of

others), observation, and analysis of records.

The survey strategy can be utilized for all four of Kirkpatrick's levels of evaluation: Self-reports to assess learners' reactions to training (often referred to as "smile tests" or "happy sheets"); self-reports, reports of others, and observations to provide evidence of learning and/or behavior change; and observations and analysis of records to assess both behavior change and results level impact on company goals.

One can insure increased validity in a survey strategy by employing certain tactics set forth by various authors in this review and elsewhere. These include proper instrument design, involving awareness of questionnaire constraints (Brandenburg, 1982a; Connolly, 1983; Maher & Kur, 1983), interview preparation and conduction techniques (Bradley, 1983; Grenough & Dixon, 1982; Salinger & Deming, 1982), and use of an adequate sample size, which is helpful for other strategies also (Smith, 1980b).

Eighty percent of the articles reviewed refer to the survey strategy as a way of obtaining evaluation evidence or information. Self-reports and reports of others are the most frequently discussed source, most often used to assess behavior change (see the Master Data Chart in the Appendix).

While the use of questionnaires and interviews is very common, there is some controversy in the reviewed articles regarding the validity of these self-reports and reports of others, particularly as the primary data source for determining behavior change (Brown, 1982; Zenger & Hargis, 1982). Brown laments the fact that verbal report data were the only measure of behavior change in an ASTD series of papers on the impact of management training.

On the other hand, Connolly (1983) devotes a whole article to the virtues of "participant evaluation" believing it is one of the best ways of determining the effect of training on work habits. She relates how validity is increased by several factors: assuring anonymity, having not only trainees respond, but their superiors

and subordinates as well, and by having an independent third-party evaluator collect the data systematically.

Mezoff (1981) states that while findings show most program evaluators and researchers consider self-reports to be the least accurate and reliable forms of measuring participant change (Howard, Maxwell, Weiner, Boynton & Rooney, 1980), such beliefs have been disputed recently by these same researchers.

One method for improving the accuracy of self-reports for determining perceived change is mentioned by several authors (Benjamin, 1982; Mezoff, 1981; Powers, 1983; Preziosi & Legg, 1983). They promote the use of "retrospective pre-post" or "post-then" self-report measures rather than pre-post self-reports, to avoid response-shift bias. This bias occurs because the trainee has a different point of reference before training (he/she may perceive of having more competence than is really the case) than afterwards (when looking back), which results in minimizing the reported impact of training on behavior change.

While the accuracy of self-reports related to change may be improved with retrospective measures, they remain perceptions or opinions. A safeguard against their inherent subjectivity is to triangulate the self-reports with reports of others (Connolly, 1983) and the use of other survey measures such as analysis of employee records and/or direct observation to help confirm the self-report results. Observation is more reliable than verbal reports but, as stated earlier, very costly and time-consuming (Cornwell, 1980b; Brown, 1982).

The survey strategy's use of records as evidence of training's cost effectiveness and/or investment return is addressed with varying emphasis by a number of articles reviewed (Barta, 1982; Brown, 1980; Dobbs, 1980; Hawes, Hutchens, & Crittenden, 1981; Head & Buchanan, 1981; Kearsley & Compton, 1981; Training Today, 1980; Salinger & Deming, 1982; Shipp, 1980; Weinstein, 1981). Requests from management to provide evidence of the financial worth of their function are

surely pressing hard at the heels of training managers. A variety of approaches are offered in the above articles, some of which deal with cost-benefit estimates for deciding if a program should be initiated or when hard data is unavailable.

In the Fifth Encyclopedia of Educational Research, Swanson (1982) writes that there are few empirical studies in the research literature regarding training cost-analysis. He goes on to say that in accepted practice, "informed and systematic estimates" (p. 865) are made more often than the use of methodologically rigorous practices. This seems to parallel the previous discussion regarding the practical aspects of evaluation versus the less applicable experimental research designs for measuring the overall effects of training. The key terms appear to be "informed" and "systematic" which connote confirmability and accuracy. In program evaluation, these latter terms can be used sometimes instead of reliability and validity (Welch, Note 1), although for cost-effectiveness analysis, specifically planned for and collected hard data are always preferable to estimates (Shipp, 1980).

With all of the above survey strategy tactics, however, one cannot say with certainty that the intervention (training) caused the behavior or attitude change; a more rigorous, controlled design would be needed. Surveys can, however, provide data regarding the relevance of training to particular work assignments (Smith, 1980b).

#4 - Personnel or client assessment. Twenty-four percent of the reviewed articles refer to personnel assessment as an evaluation strategy. While tests for measuring knowledge are economic to administer, they do not measure behavior change. In order to evaluate behavior change, several assessment techniques are proposed in this literature. These include the use of assessment centers (Byham, 1982), having trainees retake the final performance exam later on (Cornwell, 1980b), and performance analysis to assess transfer of learned skills to the job (Hahne, 1981; Salinger & Deming, 1982). Sources are available for improving the validity of

tests, also. Resenberg & Smitley (1983) present a "refresher course" on test construction in their recent article, "Constructing Tests That Work."

#5 - Systematic expert judgment. This strategy refers to evaluation by knowledgeable persons regarding the training program as a whole, rather than the performance of individual participants. Coffman (1979, 1980) describes a method of using groups of "key clients" (usually the immediate bosses or supervisors of trainees) to gather and consolidate data regarding the program's strengths and weaknesses and overall results followed by the design of an action plan. In their article on cost-benefit analysis, Head and Buchanan (1981) promote the use of accountants in developing cost of training estimates.

#6 - Case study. Three of the articles reviewed briefly mention using a case study approach as an investigative method for evaluation of training programs (Brinkerhoff, 1981; Cook, 1980; Dopyera & Pitone, 1983), and one article is completely devoted to this strategy (Brinkerhoff, 1983). This method deals with qualitative data derived from "a wide range of methodologies that treat multiple variables and processes (either formally or informally), entail direct involvement of the evaluator (perhaps even in a participant role), and are concerned with one functioning unit at a time" (Anderson & Ball, 1978, p. 61).

Case study strategy is somewhat akin to what organization development specialists do initially in action research: through in-depth interviews, open-ended questionnaires, a lot of participant observation and just plan "nosing around", an ethnographic analysis is performed on some aspect of the program. The evaluator then proceeds to "paint a picture" in narrative form using this aggregate of qualitative data. This strategy is especially useful in program implementation evaluation and also for discovering any ultimate side effects of the program which may be overlooked through a strict goal achievement framework (Welch, Note 1).

As with any data collection design, there are contextual constraints to

consider and a variety of methods and tactics to use with the case study strategy such as interviewing and collecting work samples and records. Whereas most quantitative design processes are planned for in advance with the criterion available clearly pre-determined, qualitative strategies are open to whatever information/ evidence is uncovered during the process. Confirmation of findings is the key: various examples of an effect or impact need to be obtained.

In summing up this section on strategies and designs, the fact that evaluation is a highly political, value-laden activity deserves recognition. (Experimental design is supposedly a value-free model, but it is a myth that any form of behavioral science, evaluation in particular, can be value-free.) Evaluation is made up of value judgments, albeit for specific purposes such as decision making. Indeed, "value" is the root of the word, "evaluation."

The quest for "hard", quantitative data in the training and development literature without acknowledging the values aspect of evaluation reflects what Peters and Waterman (1982) state in their popular book, In Search of Excellence: "The rational model causes us to denigrate the importance of values" (p. 51).

The important point, then, for training specialists is to be aware of the relative trade-offs of the various evaluation strategies/designs in terms of rigor and practicality. The methodology must suit the evaluation's purpose, the specific questions to be answered, and the context in which the evaluation is taking place.

Tactics

Tied very closely to evaluation designs and strategies are the tactics, or "hows", for carrying them out. Considerable reference to tactics was made in the discussion of designs/strategies.

In planning an evaluation, the evaluator selects the appropriate tactics, both technical and managerial, for implementing a specific design or strategy. For example, an observation checklist (instrument) may be used in an experiment, syste-

matic "expert" judgment, survey, and case study (Anderson & Ball, 1978). As stated before, several different sources of information are recommended for an evaluation study in order to provide greater understanding of the issue and for confirmation purposes (Brinkerhoff, 1981).

Table 5 lists evaluation tactics (Welch, Note 1) and the frequency with which they are mentioned in these articles.

TABLE 5

FREQUENCY OF REFERRAL TO COMMON EVALUATION TACTICS BY THE 41 REVIEWED ARTICLES

Tactics	No. of Articles	Percent of Articles
Technical:		
1. Evaluation structure/design	17	41
2. Instrumentation	34	83
3. Sampling Plan	7	17
4. Interview protocols	5	12
5. Factors which address accuracy standards	13	32
6. Sources of information	40	98
7. Methods of collection	32	78
8. Ways to analyze and interpret data	23	56
Managerial:		
A. Who's involved	13	32
B. Timeline	6	15
C. Budget	3	7
D. Constraints	12	29

The technical tactics of instrumentation, sources of information, and methods of collection were referred to by the majority of the authors. These reflect the basic "how to do it" functions of an evaluator, along with ways to analyze and interpret data which are mentioned in 23 of the articles. Brinkerhoff, et al. (1983) provide helpful guidelines and procedures for both quantitative and qualitative analysis and the interpretation of analysis results. Forty-one percent of the articles presented some thoughts on the structure of an evaluation. This crucial tactic is closely tied to planning, which is discussed later.

Managerial tactics are also critical for an effective evaluation. These were addressed infrequently in the articles reviewed, especially tactics dealing with timeline and budget. Connolly (1983) and Dopyera & Pitone (1983) present the best overall evaluation management guides, followed by Salinger and Deming (1982) and Smith (1980a). It should be pointed out that while constraints are referred to in 51% of the articles (see the Master Data Sheet in the Appendix), the management of constraints is discussed in only 29% of the articles. Clearly, skill in the technical tactics for carrying out an evaluation is necessary, but as with the overall training function, success is enhanced by adequate management of the process.

Reporting/Utilizing Evaluation Results

Thirteen of the articles (32%) refer to the reporting component of evaluation. Perhaps the minimal attention to reporting is related to the low number of authors who address the necessary linkage between audience and question(s).

Reporting or communicating evaluation results, while often viewed as a management function, is a major component of evaluation. No evaluation is complete until it has been reported and, in the opinion of many, acted upon (Connolly & Porter, 1980; Putman, 1980). "If management doesn't know it, then it's the same as not having done it" (King & Roth, 1981, p. 66).

The evaluation report should be "easy to understand" (Cook, 1980, p. 7), and the results should be reported attractively along with a catchy title (Vickers, 1982). "Which would you rather read: an innocuous paper entitled 'Evaluation Results,' or 'Profits Soar After Training Boosts Production 20%'" asks Vickers (p. 84). He wisely adds that the person responsible for the evaluation should be present when management receives the report for explanation purposes. Being timely with evaluation results is extremely important. The most sophisticated evaluation in the world will be wasted unless the findings reach the decision maker in time to make the decision (Smith, 1980a, 1980b).

Grenough and Dixon (1980), in describing their reporting procedures as consultants, mention presenting conclusions and recommendations along with the evaluation findings. It is important to recognize the difference between conclusions and recommendations. The agreed-upon services of the evaluator, along with the components of evaluation question(s) and purpose, should indicate if judgments and/or recommendations are expected. The recommendations can be the most powerful part of the evaluation report (Morris & Fitz-Gibbon, 1978).

Results need to be presented in a way that facilitates use. This may involve educating the consumer(s) in their use (Brinkerhoff, 1981; Putman, 1980). To help insure utilization of evaluation reports, the involvement of managers (decision makers) early on in evaluation design and implementation is emphasized in the program evaluation field (Anderson & Ball, 1978) along with some authors in this review (Brinkerhoff, 1981; Coffman, 1980; Dobbs, 1980; Dopyera & Pitone, 1983; Hahne, 1981; King & Roth, 1981; Smith, 1980). The term, "credibility," (of the training function) permeates many of these articles. Results which the primary audience can use when reporting to his/her superior(s) are especially valuable (King & Roth, 1981).

Planning the Evaluation

A third of the articles mention planning, the references ranging from single statements to a matrix guide devised by Brethower and Rummler (1976) and described by Smith (1980) which covers methodology in-depth. Connolly (1983) is the other author who describes the planning and design process in detail.

Planning and designing is a critical phase in summative evaluation. It involves being/becoming familiar with the program and the context in which the evaluation will take place, identifying the constraints on the evaluation, determining the audience's concerns and the evaluation's purpose, selecting the methods to carry it out, and finally, designing the entire process as part of the evaluation proposal.

During the planning and design phase, the needed data sources are spelled out, along with the appropriate methods for obtaining and analyzing the evidence. Measures such as pre-tests and "hard" cost-effectiveness data can then be obtained rather than having to settle for cost estimates and/or retrospective measures due to lack of planning.

Undergoing the above planning and design steps does not guarantee utilization of results, however. Under-utilization of evaluation results may be due to inappropriate designs which emphasize "rigor" of method instead of designs which acknowledge real world policymaking needs (Connolly & Porter, 1980).

Thus, credibility is achieved, in part, through visible and meaningful planfulness. All participants, including decision makers, need to be involved in planning. This helps insure ownership of an evaluation relevant to decision making needs and also the beginning of communication and trust for the entire evaluation process.

Dopyera and Pitone (1983) address evaluation planning in depth. They discuss eight decision areas to be addressed in planning:

- Should an evaluation be done?
- What is the purpose of the evaluation?
- What will be measured?

- How comprehensive will the evaluation be?
- Who has the authority and responsibility?
- From where will the data come?
- How will the data be collected and compiled?
- How will the data be analyzed and reported? (p. 66)

The authors view these decision points as planning aids which will "encourage more systematic thinking about the role of evaluation as a training management tool" (p. 71). This is one of the few articles which addressed "meta-evaluation", or the concept of evaluating the evaluation, a process which should be planned for as much as the evaluation itself.

The planning and design of a program's total evaluation (including needs assessment, formative evaluation, and summative evaluation) should take place when the training program itself is being developed (Minick & Medlin, 1983). If the training manager involves the decision makers then, as some of the authors in this review suggest, last minute demands to justify cost and prove effectiveness can be avoided (Dopyera & Pitone, 1983).

Following through with a "multi-stage" evaluation approach helps to insure that the various assumptions regarding the training program are tested. For example, Minick and Medlin's (1983) "anticipatory evaluations" are meant to insure valid training program goals (based on organizational goals). Pre-measures of the organizational goal function (such as production level) are recommended for a baseline. Following implementation (formative) evaluation, learning and behavior change can be assessed using a pre-post, control group design. Of those who demonstrate learning and behavior change as a result of the training, evaluation is continued in order to assess the impact of that behavior on the organizational goals referred to in the anticipatory evaluation. At this point, a pre-post, control group design may be used and/or the success-case method described by Brinkerhoff (1983).

Evaluator's Role and Qualifications

Given the above components of systematic summative evaluation, the role and qualifications of those responsible for overseeing their implementation should be addressed.

Role

A general view of program evaluation is that it "provides services to decision makers" (Anderson & Ball, 1978, p. 6). The exact nature of these services, including evaluator role and responsibilities, needs to be spelled out and agreed upon by the evaluator and the audience(s).

It is apparent from this review that when summative evaluation of training and development programs takes place, this service is primarily performed by in-house training and development personnel. Only three of the 14 articles referring specifically to evaluator role and/or qualifications discuss using an external evaluator (Connolly, 1983; Dopyera & Pitone, 1983; Smith, 1980a). Connolly states, "Ideally, the program evaluator should be an individual who will not be affected at all by the results of the study. This person could be an independent consultant or someone from within the organization who is able to be objective because he or she has no vested interest in the outcome of the study" (p. 92). Smith recommends getting an "evaluation specialist" who doesn't necessarily have to conduct the evaluation, but can "supervise or train the evaluators or simply advise your organization as a consultant" (p. 75). Thus, the r ons for having an outsider are different for each of these authors; one is concerned with objectivity, and the other wants to make sure the job is planned and carried out correctly. Both aims are important.

In the program evaluation field, evaluators are discouraged from conducting summative evaluations of programs in which they are otherwise involved. Commenting on the need for independence, Wholey and others (1972) point out, "No program manager should be expected to evaluate the worth of his program. . . nor

should a member of the [program] manager's staff be put in the position of having to criticize his boss" (p. 23). Anderson and Ball (1978) state, "Independence in fact as well as on paper is an important posture to maintain whenever the credibility of a program is at stake" (p. 7).

The possibility of bias, or conflict of interest, when summative evaluation is performed by training and development personnel does not appear to be a general concern in the training and development field, however. Only two articles consider this issue (Connolly, 1983; Dopyera & Pitone, 1983).

Factors to weigh regarding the use of an external evaluator include money, internal personnel, credibility, communications, and objectivity. An outside evaluator should have an understanding of the training and development function, organizations, and organizational behavior, as this helps to insure the collection of valid and reliable data (Argyris, 1968). When considering use of internal training staff, expertise and the cost effectiveness of their evaluation time need to be considered (Dopyera & Pitone, 1983).

Some of the reviewed articles which address role and/or qualifications of the evaluator emphasize establishing credibility of the training and development function through the evaluation behaviors of the training and development specialist (Hahne, 1981; King & Roth, 1981; Zenger and Blitzer, 1981). The process of developing credibility, however, is multi-faceted. Much depends on what the primary audience values. If the training and development function is already credible in the eyes of the primary audience, then perhaps its own summative evaluation reports may also be viewed as credible. At any rate, the credibility of the evaluator needs to be established beforehand. Dobbs (1980) sums it up well: "When you need it [credibility] the most, it's probably too late to start building it" (p. 14).

Ethical responsibilities of the evaluator are also very important and tie in

closely with the credibility issue. Professional program evaluators have recently developed standards to follow when conducting evaluations. (ERS Standards Committee, 1982; The Joint Committee on Standards, 1981).

One suggestion for handling and/or preventing problems associated with summative evaluation, such as those related to evaluator role, is to establish an evaluation advisory board (Anderson & Ball, 1978). Such advisory groups, often comprised of senior management representatives, are common within business for setting policies for a particular function and overseeing their implementation. The board could provide advisement for summative evaluation efforts of all departments, not only training and development, and contract for external evaluators when deemed necessary.

Qualifications

The training and development specialist obviously needs to be knowledgeable about evaluation and be able to conduct certain types of evaluations, particularly needs assessments and formative evaluations for program development and improvement purposes. However, he/she cannot be expected to possess all the skills of a qualified professional program evaluator, whose competencies fill six pages in a recent publication (Udinsky, Osterlind, & Lynch, 1981).

It is noteworthy that the American Society for Training and Development has completed a competency study for its field which describes competency in program evaluation as follows:

Knowledge of:

- Levels of program evaluation, including evaluation focusing on:
  - participants' reactions
  - knowledge development
  - skills development
  - values/attitude change
  - application

- business results

Skill in:

- Developing evaluation designs that economically, reliably, and validly gather data on program effectiveness.
- Setting up appropriate evaluation conditions.
- Conducting the evaluation (McLagan, 1982).

Various skills desired of program evaluators are mentioned in the 41 articles and incorporated in the evaluation component sections discussed earlier in this paper. The general message one is given, however, is that training and development specialists for the most part are not very skilled in systematically designing and conducting effective summative evaluations, including those aimed at fiscal accountability (Bakken & Berstein, 1982; Cook, 1980; Head & Buchanan, 1981; Weinstein, 1982).

While awareness of the need for improvement in this process appears to be high (Olivas & Iman, 1983), the avenues by which most training specialists may attempt to broaden their evaluation knowledge and skills may be questioned. The training and development literature does not provide a cohesive, system approach. One would suspect that sharing ideas with other training specialists would only perpetuate an aimless search for "the" way to evaluate. It would not provide the necessary background theory and knowledge to make appropriate evaluation decisions (including those pertaining to role), given specific situations.

Even some of the opportunities to upgrade one's skills within the training and development field are questionable for becoming more qualified in evaluation. According to a recent article regarding the validity of "train the trainer" programs, quality and depth vary considerably from program to program, and evaluation was addressed in less than half of those programs studied (DiPaolo & Patterson, 1983).

Training and development professional organization seminars which stress a

system approach to evaluation, along with university and college courses in program evaluation, would benefit training specialists. The latter would provide and promote in-depth study and exposure to the program evaluation literature, along with opportunities for practice through course projects. The former would address the specific aspects of designing and conducting systematic evaluation peculiar to private sector situations. Through a broadened view of program evaluation, including state-of-the-art issues and methodology, training specialists would be able to integrate evaluation into a total training and development system.

Interpersonal and public relations skills are discussed in the program evaluation/research literature, in program evaluation training (Anderson & Ball, 1978; Argyris, 1968; Welch, Note 1), and in some of these reviewed articles, as being very important. While the evaluator may have a strong spirit of inquiry, the individuals with the needed information/evidence for the evaluation (including decision makers) may require considerable time with the evaluator to develop trust and openness. It is essential to nourish this trust and openness throughout the entire evaluation process by keeping all participants informed and audiences involved with each other at all times (Argyris, 1968; Welch, Note 1).

Little or no mention is made in the training and development literature regarding professional and ethical sensitivity as a necessary qualification for evaluators (except for Brown, 1980, who speaks against withholding certain types of programs, such as safety training, from one group of employees in order to utilize a control group).

Nickols (1983), in his rejoinder to an article predicting the state of performance and instruction in the year 2000, states that politics is the name of the game when it comes to measuring the value of training. Likewise, Dopyera and Pitone (1983) contend that politics is part of evaluation reporting. They state, "What is played up or played down in reporting is a political issue that cannot be

avoided. In large organizations, different people will want different results" (p. 71). Given that this is the position in which some internal evaluators are placed, it is not surprising that their summative evaluation efforts may be viewed by them as a waste of time. The competent external program evaluator, in upholding his/her profession's standards, strives for objective reporting while being aware of the political context in which he/she operates. Evaluators also need to be aware of their own personal values and how those values may influence their judgments in an evaluation (Anderson & Ball, 1978; ERS Standard Committee, 1982; The Joint Committee on Standards, 1981; Welch, Note 1).

Conclusion

This monograph has reviewed the literature in training and development of the four years from 1980 through 1983 as it pertains to summative evaluation.

All of the standard components of program evaluation are addressed either directly or indirectly in these 41 articles, in varying degree and depth. Strategies/designs and tactics are covered in all but one article (King & Roth, 1981), while a small percentage (29%) address the reporting of evaluation results. Few authors link the component of specific audience(s) to evaluation question(s), an especially crucial aspect of program evaluation. Only 34% discuss the planning and designing of an evaluation, and only 37% refer to the role and/or qualifications of the evaluator.

The general consensus of the authors is that most summative evaluation of training and development programs is not conducted effectively at the present time. It may be hypothesized that more attention needs to be given to the components of summative evaluation used as the basis for this review and to upgrading the evaluation competencies of training specialists.

Reference Notes

1. Welch, W. Principles and methods of evaluation and Evaluation methods and techniques. Courses conducted at the University of Minnesota, Minneapolis, MN. 1981-1982.

References

Anderson, S. B. & Ball, S. (1978). The profession and practice of program evaluation. San Francisco: Jossey-Bass.

Argyris, C. (1962). Interpersonal competence and organizational effectiveness. Homewood, IL: Irwin.

Argyris, C. (1968). Some unintended consequences of rigorous research. Psychological Bulletin, 70(3), 185-197.

*Bakken, D. & Bernstein, A. L. (1982). A systematic approach to evaluation. Training and Development Journal, 36(8), 44-51.

*Barta, T. (1982). Methods to determine return on the training investment. NSPI Journal, 21(1), 16-17; 24.

*Benjamin, E. R. (1982). Using the 'post-then' method of evaluation. Training/HRD, 19(11), 72.

*Blakeslee, G. S. (1982). Evaluating a communications training program. Training and Development Journal, 36(11), 84-89.

Borg, W. R. & Gall, M. D. (1979). Educational research. New York: Longman.

Bradley, J. (1983). How to interview for information. Training, 20(4), 59-62.

*Brandenburg, D. C. (1982). Some improvements for the ubiquitous evaluation questionnaire. NSPI Journal, 21(8), 9-11.

Brandenburg, D. C. (1982). Training evaluation: What's the current status? Training and Development Journal, 36(5), 14-22.

Brethower, K. S. & Rummler, G. A. (1976). Evaluating training. Improving Human Performance Quarterly, 5(3-4), 103-120. Reprinted in Training and Development Journal, 1979, 33(5), 14-22.

*Brinkerhoff, R. O. (1981). Making evaluation more useful. Training and Development Journal, 35(12), 66-70.

*Brinkerhoff, R. O. (1983) The success case: A low-cost, high-yield evaluation. Training and Development Journal, 37(8), 58-59; 61.

Brinkerhoff, R. O., Brethower, D. M., Hluchyj, T. & Nowakowski, J. R. (1983). Program evaluation: A practitioner's guide for trainers and educators. Boston: Kluwer-Nijhoff.

*Brown, M. G. (1980). Evaluating training via multiple baseline designs. Training and Development Journal, 34(10), 11-16.

*Brown, M. G. (1982). Translating managerial behavior into measurable accomplishments. NSPI Journal, 21(1), 13-15.

*Byham, W. C. (1982). How assessment centers are used to evaluate training's effectiveness. Training/HRD, 19(2), 32-38.

Campbell, D. T. & Stanley, J. C. (1966). Experimental and quasi-experimental designs for research. Chicago: Rand McNally.

Campbell, J. P. (1971). Personnel training and development. Annual Review of Psychology, 22, 565-602.

Catalanello, R. F. & Kirkpatrick, D. L. (1968). Evaluating training programs--The state of the art. Training and Development Journal, 22(5), 2-9.

*Clement, R. W. & Aranda, E. K. (1982). Evaluating management training: A contingency approach. Training and Development Journal, 36(8), 39-43.

Coffman, L. (1979). An easy way to effectively evaluate program results. Training and Development Journal, 33(8), 28-32.

*Coffman, L. (1980). Successful training program evaluation. Training and Development Journal, 34(10), 84-87.

*Connolly, S. (1983). Participant evaluation: Finding out how well training worked. Training and Development Journal, 37(10), 92-96.

Connolly, T., & Porter, A. L. (1980). A user-focused model for the utilization of evaluation. Evaluation and Program Planning, 3, 131-140.

*Cook, M. H. (1980). Training isn't a cost--it's a blue-chip investment. Training and Development Journal, 34(2), 4-7.

*Cornwell, J. B. (1980). Quality assurance and control in training. Training/HRD, 17(1), 33-34.

*Cornwell, J. B. (1980). Justify your training efforts with on-the-job performance. Training/HRD, 17(8), 34-36.

DiPaolo, A. J., & Patterson, A. C. (1983). Selecting a training program for new trainers. Training and Development Journal, 37(1), 96-98; 100-101.

*Dobbs, J. H. (1980). Building training department credibility. Training and Development Journal, 34(3), 14-21.

*Dopyera, J., & Pitone, L. (1983). Decision points in planning the evaluation of training. Training and Development Journal, 37(5), 66-71.

ERS Standards Committee. (1982). Evaluation research society standards for evaluation practice. In P. H. Rossi (Ed.), Standards for Evaluation Practice, New Directions for Program Evaluation, No. 15, 7-19.

*Grenough, J. L., & Dixon, R. W. (1982). Using 'utilization' to measure training results. Training/HRD, 19(2), 40-42.

*Hahne, G. (1981). Creating credibility for your sales training. Training and Development Journal, 35(11), 36-38.

*Hanson, M. C. (1982). Career/life planning workshops as career services in organizations--Are they working? Training and Development Journal, 36(2), 58-63.

*Hawes, J. M., Hutchens, S. P., & Crittenden, W. F. (1982). Evaluating corporate sales training programs. Training and Development Journal, 36(11), 44-46; 48.

*Head, G. E., & Buchanan, C. C. (1981). Cost/benefit analysis of training: A foundation for change. NSPI Journal, 20(9), 25-27.

*Holoviak, S. J. (1982). The impact of training on company productivity levels. NSPI Journal, 21(5), 6-8.

Howard, G. S., Maxwell, S. E., Weiner, R. L., Boyton, K. S., & Rooney, W. M. (1980). Is a behavioral measure the best estimate of behavioral parameter? Perhaps not. Applied Psychological Measurement, 4(3), 292-311.

The Joint Committee on Standards for Educational Evaluation. (1981). Standards for Evaluations of Educational Programs, Projects, and Materials. New York: McGraw-Hill.

*Kearsley, G., & Compton, T. (1981). Assessing costs, benefits and productivity in training systems. Training and Development Journal, 35(1), 52-61.

*King, W. M., & Roth, R. A. (1981). Impact-directed training. Training and Development Journal, 35(11), 61-63; 65-69.

Kirkpatrick, D. L. (1960). Techniques for evaluating training programs. Training and Development Journal, 12(6). Reprinted in Training and Development Journal, 1979, 33(6), 78-92.

Lindholm, T. R. (1953). Supervisory training and employee attitudes. Journal of the ASTD, November-December.

Maher, J. H., Jr., & Kur, C. E. (1983). Constructing good questionnaires. Training and Development Journal, 37(6), 100; 103-108; 110.

*McAfee, R. B. (1982). Using performance appraisals to enhance training programs. Personnel Administrator, 28(11), 31-34.

McLagan, P. A. (1982). Models for excellence: The conclusions and recommendations of the ASTD competency study. Washington, D.C.: ASTD Press.

*Meier, T. K., & Pulichene, J. P. (1980). Evaluating the effectiveness of assertiveness training. Training and Development Journal, 34(2), 66-68.

*Mezoff, B. (1981). Pre-then-post testing: A tool to improve the accuracy of management training program evaluation. NSPI Journal, 20(8), 10-11; 16.

Minick, R. D., & Medlin, S. M. (1983). Anticipatory evaluations in HRD programming. Training and Development Journal, 37(5), 89-94.

Mitroff, I. I., & Featheringham, T. R. (1974). On systematic problem solving and the error of the third kind. Behavioral Science, 19, 383-393.

Morris, L. L., & Fitz-Gibbon, C. T. (1978). How to present an evaluation report. Beverly Hills: Sage.

Nickols, F. W. (1983). Performance & instruction in the year 2000: A rejoinder. Performance & Instruction Journal, 22(8), 8-9.

Olivas, L., & Inman, T. (1983). What concerns today's trainers? Training and Development Journal, 37(7), 62; 64.

Peters, T. J., & Waterman, R. H., Jr. (1982). In search of excellence. New York: Harper & Row.

*Powers, E. (1983). Step-by-step to the design of sales training evaluation. NSPI Journal, 22(2), 23-25; 38.

*Preziosi, R. C., & Legg, L. M. (1983). Add 'then' testing to prove training's effectiveness. Training, 20(5), 48-49.

*Putman, A. O. (1980). Pragmatic evaluation. Training and Development Journal, 34(10), 36-40.

Rosenberg, M. J., & Smitley, W. (1983). Constructing tests that work. Training, 20(9), 41-42; 44; 48.

*Salinger, R. D., & Deming, B. S. (1982). Practical strategies for evaluating training. Training and Development Journal, 36(8), 20-26; 28-29.

*Shipp, T. (1980). To survive the budget inquisition, prove your training makes dollars and sense. Training/HRD, 17(11), 23; 26; 28.

Smeltzer, L. R. (1979). Do you really evaluate, or just talk about it? Training, 17(8), 6-8.

*Smith, M. E. (1980a). Evaluating training operations and programs. Training and Development Journal, 34(10), 70-78.

*Smith, M. E. (1980b). How big a sample do I need for my evaluation? NSPI Journal, 19(10), 3-7; 10.

Swanson, R. A. (1982). Industrial training. Encyclopedia of Educational Research--Fifth Edition, 2, 864-870.

Swanson, R. A. (Ed.). (1983). Performance and Instruction Journal, 22(5).

*Training today. Training/HRD, 1980, 17(1), 6.

Udinsky, B. F., Osterlind, S. J., & Lynch, S. W. (1981). Evaluation resource handbook: Gathering, analyzing, reporting data. San Diego: EdITS.

*Vickers, T. W. (1982). You know your value . . . but can you prove it? Training/HRD, 19(8), 84.

Webster's New Collegiate Dictionary. (1981). Springfield, MA.: G. & C. Merriam.

*Weinstein, L. M. (1982). Collecting training cost data. Training and Development Journal, 36(8), 31-34.

Weiss, R. S., & Rein, M. (1970). The evaluation of broad-aim programs: experimental design, its difficulties, and an alternative. Administrative Science Quarterly, 15, 97-109.

Wholey, J. S., Duffy, H. G., Fukumoto, J. S., Scanlon, J. W., Berlin, M. A., Copeland, W. C., & Zelinsky, J. G. (1972). Proper organizational relationships. In C. H. Weiss (Ed.), Evaluating action programs: Readings in social action and education. Boston: Allyn & Bacon, 118-122.

*Zenger, J. H., & Blitzer, R. J. (1981). How training managers become corporate heroes/heroines. Training/HRD, 18(11), 22; 27.

*Zenger, J. H., & Hargis, K. (1982). Assessing training results: It's time to take the plunge! Training and Development Journal, 36(1), 11-16.

*Reviewed articles

APPENDIX

MASTER DATA CHART

MASTER DATA CHART

REVIEWED ARTICLES	PROGRAM EVALUATION COMPONENTS ADDRESSSED									
	AUDIENCE(S)	QUESTION(S)	PURPOSE[1]	CONTEXT & CONSTRAINTS	METHODS & PROCEDURES: FRAMEWORK/ MODEL[2]	METHODS & PROCEDURES: STRATEGIES/ DESIGNS[3]	METHODS & PROCEDURES: TACTICS[4]	REPORTING UTILIZING EVALUATION RESULTS	PLANNING	EVALUATOR ROLE & QUALIFICATIONS
1. BAKKEN & BERNSTEIN (1982)	X	X	2,3		5,7	2,3,4	1,2,5, 6,7		X	X
2. BARTA (1982)	X	X	3		8	3	6			
3. BENJAMIN (1982)		X	2,3		7	3	2,5,6,7,8			
4. BLAKESLEE (1982)		X	2,3	X	N/A	3	2,6,7,8		X	
5. BRANDENBURG (1982)	X	X	2,3		5	3	2,6,8			
6. BRINKERHOFF (1981)	X	X	1,2,3	X	5	2,3,4, 5,6	1,2,5, 6,7	X	X	
7. BRINKERHOFF (1983)	X	X	2,3	X	5	6	1,2,3,4, 5,6,7,8,A	X		
8. BROWN (1980)	X	X	3	X	7,8	2,3	2,6,7,8			
BROWN (1982)	X		3	X	7,8	2,3	1,2,6,7			
10. BYHAM (1982)		X	3		7	4	2,6,7,A			
11. CLEMENT & ARANDA (1982		X	2,3,4		8	2	6,7		X	

REVIEWED ARTICLES	PROGRAM EVALUATION COMPONENTS ADDRESSSED									
	AUDIENCE(S)	QUESTION(S)	PURPOSE[1]	CONTEXT & CONSTRAINTS	METHODS & PROCEDURES: FRAMEWORK/ MODEL[2]	METHODS & PROCEDURES: STRATEGIES/ DESIGNS[3]	METHODS & PROCEDURES: TACTICS[4]	REPORTING UTILIZING EVALUATION RESULTS	PLANNING	EVALUATOR ROLE & QUALIFICATIONS
12. COFFMAN (1980)	X	X	1,3	X	7	4,5	1,2,4,6 7,A,D		X	X
13. CONNOLLY (1983)	X	X	2,3	X	5	3	1,2,3,4, 5,6,7, A,B,C,D		X	X
14. COOK (1980)	X	X	2,3,4		7,8	2,3,6	1,2,6,7,8	X	X	X
15. CORNWELL (1980a)		X	2,3		7	3	2,6,8			
16. CORNWELL (1980b)		X	2,3	X	7,8	2,3,4	2,6,7			
17. DOBBS (1980)	X	X	3,4	X	5,7	3	1,2,4, 6,A,D	X	X	X
18. DOPYERA & PITONE (1983)	X		1,2,3,4	X	7	1,3,4,6	1,2,3,5, 6,7,8,A, B,C,D	X	X	X
19. GRENOUGH & DIXON (1982)	X	X	2,3	X	8	3	1,2,6, 7,D	X		X
20. HAHNE (1981)	X	X	4		7	3,4	2,6,7,B	X		X
21. HANSON (1982)		X	3,4		7,8	2,3	1,2,6, 7,8			
22. HAWES, ET AL. (1982)			2,3	X	7	2,3	2,6,7,8,D			

REVIEWED ARTICLES	PROGRAM EVALUATION COMPONENTS ADDRESSSED									
	AUDIENCE(S)	QUESTION(S)	PURPOSE[1]	CONTEXT & CONSTRAINTS	METHODS & PROCEDURES: FRAMEWORK/ MODEL[2]	METHODS & PROCEDURES: STRATEGIES/ DESIGNS[3]	METHODS & PROCEDURES: TACTICS[4]	REPORTING UTILIZING EVALUATION RESULTS	PLANNING	EVALUATOR ROLE & QUALIFICATIONS
23. HEAD & BUCHANAN (1981)	X	X	3		8	3,5	1,6,8,A		X	X
24. HOLOVIAK (1982)		X	3		7	3	2,6,8			
25. KEARSLEY & COMPTON (1981)			3		8	3	1,2,6,8	X		
26. KING & ROTH (1981)	X	X	4		7			X		X
27. MCAFEE (1982)	X	X	1,3	X	5,7	2	2,6,7,D			
28. MEIER & PULICHENE (1980)		X	3		7	2	2,6,7,8			
29. MEZOFF (1981)		X	3	X	7	3	2,5,6, 7,8			X
30. POWERS (1983)	X	X	3,4		5,7	2,3	2,3,6, 7,8,A,			
31. PREZIOSI & LEGG (1983)	X	X	4		7	3	2,5,6,7,8			
32. PUTMAN (1980)	X	X	3	X	5,7,8	3	1,2,6, 7,8,A	X		
33. SALINGER & DEMING (1982)	X	X	1,2,3	X	7,8	2,3,4	2,3,4,5,6, 7,8,A,B,D,			

REVIEWED ARTICLES	PROGRAM EVALUATION COMPONENTS ADDRESSSED									
	AUDIENCE(S)	QUESTION(S)	PURPOSE[1]	CONTEXT & CONSTRAINTS	METHODS & PROCEDURES: FRAMEWORK/ MODEL[2]	METHODS & PROCEDURES: STRATEGIES/ DESIGNS[3]	METHODS & PROCEDURES: TACTICS[4]	REPORTING UTILIZING EVALUATION RESULTS	PLANNING	EVALUATOR ROLE & QUALIFICATIONS
34. SHIPP (1980)			3,4	X	8	3	1,5,6, 7,8,D		X	
35. SMITH (1980e)	X	X	1,2,3	X	5,7,8	3	1,2,5,6, A,B,D	X	X	X
36. SMITH (1980b)			3	X	N/A	1,2,3	3,5,6,7, 8,B,D			
37. TRAINING TODAY (Rosentreter) (1980)		X	3		7,8	1,3	2,3,6, 7,A			
38. VICKERS (1982)	X	X	3,4	X	7,8	3	6,7,A	X		X
39. WEINSTEIN (1982)	X	X	3		8	3	1,2,5, 6,7,8			X
40. ZENGER & BLITZER (1981)	X	X	3,4		7,8	3,4	2,6,7	X		X
41. ZENGER & HARGIS (1982)	X	X	3,4	X	7,8	1,2,4	2,6,7, 8,C,D		X	
TOTAL NUMBER:	26	35		21				13	14	14
PERCENT:	63%	85%		51%				32%	34%	34%

[1] Keyed to Table 2.
[2] Keyed to Table 3.
[3] Keyed to Table 4.
[4] Keyed to Table 5.

MASTER DATA CHART

REVIEWED ARTICLES	PROGRAM EVALUATION COMPONENTS ADDRESSSED									
	AUDIENCE(S)	QUESTION(S)	PURPOSE[1]	CONTEXT & CONSTRAINTS	METHODS & PROCEDURES: FRAMEWORK/ MODEL[2]	METHODS & PROCEDURES: STRATEGIES/ DESIGNS[3]	METHODS & PROCEDURES: TACTICS[4]	REPORTING UTILIZING EVALUATION RESULTS	PLANNING	EVALUATOR ROLE & QUALIFICATIONS
1. BAKKEN & BERNSTEIN (1982)	X	X	2,3		5,7	2,3,4	1,2,5, 6,7		X	X
2. BARTA (1982)	X	X	3		8	3	6			
3. BENJAMIN (1982)		X	2,3		7	3	2,5,6,7,8			
4. BLAKESLEE (1982)		X	2,3	X	N/A	3	2,6,7,8		X	
5. BRANDENBURG (1982)	X	X	2,3		5	3	2,6,8			
6. BRINKERHOFF (1981)	X	X	1,2,3	X	5	2,3,4, 5,6	1,2,5, 6,7	X	X	
7. BRINKERHOFF (1983)	X	X	2,3	X	5	6	1,2,3,4, 5,6,7,8,A	X		
8. BROWN (1980)	X	X	3	X	7,8	2,3	2,6,7,8			
9. BROWN (1982)	X		3	X	7,8	2,3	1,2,6,7			
10. BYHAM (1982)		X	3		7	4	2,6,7,A			
11. CLEMENT & ARANDA (1982		X	2,3,4		8	2	6,7		X	

REVIEWED ARTICLES	PROGRAM EVALUATION COMPONENTS ADDRESSSED									
	AUDIENCE(S)	QUESTION(S)	PURPOSE[1]	CONTEXT & CONSTRAINTS	METHODS & PROCEDURES: FRAMEWORK/ MODEL[2]	METHODS & PROCEDURES: STRATEGIES/ DESIGNS[3]	METHODS & PROCEDURES: TACTICS[4]	REPORTING UTILIZING EVALUATION RESULTS	PLANNING	EVALUATOR ROLE & QUALIFICATIONS
12. COFFMAN (1980)	X	X	1,3	X	7	4,5	1,2,4,6 7,A,D		X	X
13. CONNOLLY (1983)	X	X	2,3	X	5	3	1,2,3,4, 5,6,7, A,B,C,D		X	X
14. COOK (1980)	X	X	2,3,4		7,8	2,3,6	1,2,6,7,8	X	X	X
15. CORNWELL (1980a)		X	2,3		7	3	2,6,8			
16. CORNWELL (1980b)		X	2,3	X	7,8	2,3,4	2,6,7			
17. DOBBS (1980)	X	X	3,4	X	5,7	3	1,2,4, 6,A,D	X	X	X
18. DOPYERA & PITONE (1983)	X		1,2,3,4	X	7	1,3,4,6	1,2,3,5, 6,7,8,A, B,C,D	X	X	X
19. GRENOUGH & DIXON (1982)	X	X	2,3	X	8	3	1,2,6, 7,D	X		X
20. HAHNE (1981)	X	X	4		7	3,4	2,6,7,B	X		X
21. HANSON (1982)		X	3,4		7,8	2,3	1,2,6, 7,8			
22. HAWES, ET AL (1982)			2,3	X	7	2,3	2,6,7,8,D			

REVIEWED ARTICLES	PROGRAM EVALUATION COMPONENTS ADDRESSSED									
	AUDIENCE(S)	QUESTION(S)	PURPOSE[1]	CONTEXT & CONSTRAINTS	METHODS & PROCEDURES: FRAMEWORK/ MODEL[2]	METHODS & PROCEDURES: STRATEGIES/ DESIGNS[3]	METHODS & PROCEDURES: TACTICS[4]	REPORTING UTILIZING EVALUATION RESULTS	PLANNING	EVALUATOR ROLE & QUALIFICATIONS
23. HEAD & BUCHANAN (1981)	X	X	3		8	3,5	1,6,8,A		X	X
24. HOLOVIAK (1982)		X	3		7	3	2,6,8			
25. KEARSLEY & COMPTON (1981)			3		8	3	1,2,6,8	X		
26. KING & ROTH (1981)	X	X	4		7			X		X
27. MCAFEE (1982)	X	X	1,3	X	5,7	2	2,6,7,D			
28. MEIER & PULICHENE (1980)		X	3		7	2	2,6,7,8			
29. MEZOFF (1981)		X	3	X	7	3	2,5,6, 7,8			X
30. POWERS (1983)	X	X	3,4		5,7	2,3	2,3,6, 7,8,A,			
31. PREZIOSI & LEGG (1983)	X	X	4		7	3	2,5,6,7,8			
32. PUTMAN (1980)	X	X	3	X	5,7,8	3	1,2,6, 7,8,A	X		
33. SALINGER & DEMING (1982)	X	X	1,2,3	X	7,8	2,3,4	2,3,4,5,6, 7,8,A,B,D,			

REVIEWED ARTICLES	PROGRAM EVALUATION COMPONENTS ADDRESSSED									
	AUDIENCE(S)	QUESTION(S)	PURPOSE[1]	CONTEXT & CONSTRAINTS	METHODS & PROCEDURES: FRAMEWORK/ MODEL[2]	METHODS & PROCEDURES: STRATEGIES/ DESIGNS[3]	METHODS & PROCEDURES: TACTICS[4]	REPORTING UTILIZING EVALUATION RESULTS	PLANNING	EVALUATOR ROLE & QUALIFICATIONS
34. SHIPP (1980)			3,4	X	8	3	1,5,6, 7,8,D		X	
35. SMITH (1980a)	X	X	1,2,3	X	5,7,8	3	1,2,5,6, A,B,D	X	X	X
36. SMITH (1980b)			3	X	N/A	1,2,3	3,5,6,7, 8,B,D			
37. TRAINING TODAY (Rosentreter) (1980)		X	3		7,8	1,3	2,3,6, 7,A			
38. VICKERS (1982)	X	X	3,4	X	7,8	3	6,7,A	X		X
39. WEINSTEIN (1982)	X	X	3		8	3	1,2,5, 6,7,8			X
40. ZENGER & BLITZER (1981)	X	X	3,4		7,8	3,4	2,6,7	X		X
41. ZENGER & HARGIS (1982)	X	X	3,4	X	7,8	1,2,4	2,6,7, 8,C,D		X	
TOTAL NUMBER:	26	35		21				13	14	14
PERCENT:	63%	85%		51%				32%	34%	34%

[1] Keyed to Table 2.
[2] Keyed to Table 3.
[3] Keyed to Table 4.
[4] Keyed to Table 5.